DOCUMENT RESUME

ED 445 056 TM 031 672

AUTHOR	Impara, James C.; Plake, Barbara S.; Buckendahl, Chad W.
TITLE	The Comparability of Norm-Referenced Achievement Tests as They Align to Nebraska's Language Arts Content Standards.
PUB DATE	2000-06-00
NOTE	13p.; Paper presented at the Annual Large Scale Assessment Conference (Snowbird, UT, June 2000).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Achievement Tests; Elementary Secondary Education; *Language Arts; *Norm Referenced Tests; *State Standards; *Teachers; Validity
IDENTIFIERS	Curriculum Alignment; *Nebraska

ABSTRACT

Three separate studies were designed to provide information on the alignment of norm-referenced achievement test items in language arts to Nebraska's content standards at grades 4, 8, and 11. The information was used to evaluate the alignment of the tests to the standards as well as compare that alignment across the five tests. There were 20 teachers on the 4th-grade panel, 10 teachers on the 8th grade panel, and 10 on the 11th grade panel. Teachers from school districts across the state were selected to participate in these studies. The teachers evaluated how well items from these tests aligned to content standards in reading, writing, speaking, and listening. Results for language arts showed moderate alignment between the norm-referenced test items and standards. These studies were used as validity evidence for Nebraska's state assessment model that permits school districts to select the norm-referenced test that best meets their needs and then to supplement that test with district developed assessments to address what the norm-referenced tests do not measure. (Contains 3 tables and 10 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 445 056

The comparability of norm-referenced achievement tests as they align to Nebraska's language arts content standards.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

C. W. Buckendahl

James C. Impara

Barbara S. Plake

Chad W. Buckendahl

Buros Center for Testing

Paper presented at the Large Scale Assessment Conference at Snowbird, UT

June, 2000

TM031672

BEST COPY AVAILABLE

Abstract

Three separate studies were designed to provide information on the alignment of norm-referenced achievement test items in language arts to Nebraska's content standards at grades 4, 8, and 11. The information was used to evaluate the alignment of the tests to the standards as well as compare that alignment across the five tests. Teachers from school districts across the state were selected to participate in these studies. The teachers evaluated how well items from these tests aligned to content standards in reading, writing, speaking, and listening. Results for language arts showed moderate alignment between the norm-referenced test items and standards. These studies were used as validity evidence for Nebraska's state assessment model that permits school districts to select the norm-referenced test that best meets their needs and then supplement that test with district developed assessments to address what the norm-referenced tests do not measure.

The comparability of norm-referenced achievement tests as they align to Nebraska's language arts content standards.

Introduction

Standards-led assessment has become a concern of many states and local education agencies in recent years. An added challenge for these agencies that have adopted content standards has been the decision of an appropriate assessment (Linn & Herman, 1997). Competing interests of accountability, local control, and appropriate use of these assessments have become common themes among legislative committees, boards of education, administrators, and teachers. Many states have confronted these assessment issues and have been criticized for using a single assessment or for relying too heavily on commercially available tests that may not align to the standards. Are there data that school districts already collect that could provide information policy makers regarding student performance? For a state that has chosen to incorporate norm-referenced tests in their assessment plan, what are the limits to the information that scores from these tests can provide? How do the most prevalent tests in Nebraska compare in terms of their alignment to their language arts content standards? These are questions this research sought to answer.

The specific purpose of these studies was to collect teachers' judgments on the alignment of norm-referenced test items to Nebraska's language arts content standards and to compare those alignment judgments across tests for grades 4, 8, and 11. The studies provided validity evidence for a state assessment model that relies heavily on school districts, rather than the state measuring student performance on content standards.

Methods

These studies considered the five most prevalent norm-referenced achievement test batteries used in Nebraska school districts to measure the alignment of test items to content standards in language arts. Within each test battery, sub-tests were selected that were expected to have relevant items for inclusion in the study. Sub-tests including vocabulary, language mechanics, or reading comprehension were used for language arts.

The guiding principle behind examining the alignment of norm-referenced test items to Nebraska's content standards is that it may be possible for these tests to provide limited information to the state on student performance on content standards. There was an expectation, though, that any single achievement test alone would probably not provide adequate coverage of the standards. Similar alignment studies in other states have been conducted using teachers or coders as judges that align test items to standards given a pre-determined rubric (Webb & Smithson, 1999; Wixson, Fisk, Dutro, & McDaniel, 1999). These studies focused primarily on single grade and content areas.

Participants in these studies were teachers recruited from across the state to represent a variety of locations and sizes of school districts. There were 20 teachers for the 4th grade panel, 10 teachers on the 8th grade and 11th grade panel. Teachers were asked to participate based on the grade they taught and their specialized subject area (e.g., language arts or mathematics).

Procedures

All three standards alignment workshops were conducted during three, two-day workshops in a conference center. The same basic procedures were followed for each of the three alignment workshops and the same group facilitators were used for each of the

studies to reduce the potential biasing effects. Teachers examined each of the commercially available norm-referenced achievement test batteries in a predetermined order. For each workshop, half of the teachers examined the test batteries in the reverse order to counterbalance a potential order effect.

Prior to rating the alignment of the norm-referenced assessments to the content standards, teachers participated in a training session. Through the training session, teachers reviewed the relevant content standards and participated in a practice exercise to clarify the rating process they were to use. The practice exercise used the sample items from alternate forms of the achievement tests that were going to be rated. Teachers were allowed to discuss their ratings to gain multiple perspectives regarding interpretation of the alignment of items to state content standards. This discussion only occurred in training.

Data collection began with gathering background information about the participating teachers. This was done to verify that the panelists had experience at the grade level and content area at which they were being asked to provide judgments. Teachers were then asked to assess the extent that each test item matched the content standards. Special forms that had each sub-test name and a matrix of items by standards were prepared for this purpose. Unlike the training, teachers were not allowed to discuss their operational judgments. Teachers used the following rating criteria to judge the level of alignment of an item to standard:

High Level of Alignment = A high level of alignment indicates that the item measures the standard. This high rating means that you would be comfortable making

inferences about a student's performance on that standard by knowing their performance on this item or similar items.

Moderate Level of Alignment = A moderate level of alignment indicates that the item measures a portion of the standard. Since standards may have multiple parts, a high level of alignment may not be appropriate, but portions that are addressed by an item may warrant this level of alignment. A moderate rating means that you would be comfortable making inferences about a student's performance on that standard by knowing their performance on a collection of similar items.

Low Level of Alignment = A low level of alignment indicates that the item barely measures the standard. This level of alignment means that you probably had to stretch to find alignment between item and standard. An item that aligns with only a small portion of a standard comprised of many aspects may warrant this rating. A low rating means that you would not feel comfortable making inferences about a student's performance on the standard knowing their performance on this item or similar items.

No Alignment = No alignment means that an item does not measure any aspects of the standard. This lack of alignment means that you found no alignment between item and standard. This rating means that you would not make an inference about a student's performance on the standard knowing their performance on this item or similar items.

Teachers were instructed to consider multiple potential item-to-standard alignments, perhaps at varying levels. Therefore, any one item could be rated as having high, medium, low, or no alignment to multiple grade specific content standards. These rating categories were defined and discussed during the training component of the workshop prior to the judgment of operational items. Upon completion of these ratings,

teachers completed an evaluation form that gathered information to provide procedural validity evidence of their judgments including their satisfaction with the training and rating processes as well as the time allocated to each.

Results

The results of this study are predicated on the operational definition of adequate measurement of a standard that was used in this study. There were three criteria used to determine whether a standard was adequately measured: 1) alignment rating level, 2) teacher agreement on rating levels, and 3) number of aligned items. Item-to-standard matches that were identified by teachers at the high or moderate alignment levels met the first criterion for adequate measurement of a standard. If at least 50% of teachers (10 for 4th grade panels , 5 for 8th grade and high school panels) agreed with this high or moderate level of alignment, the second criterion for adequate measurement of a standard was met. Last, there had to be at least five items that met criteria 1) and 2).

Thus, a standard was said to be measured if a minimum of five items a) were rated at the high or moderate level of alignment by b) at least 50% of the panel. Although Webb (1997) suggests a minimum of six items for making an inference about content knowledge, we chose five items as a more lenient criterion since there was a consideration of inter-rater reliability criterion (Schmidt, 1999) included as well. These results were tabulated for the Language Arts content standards for each of the three grade levels for the five most prevalent standardized achievement tests in the state.

4th Grade Language Arts

Table 1 shows the results for the alignment of the five norm-referenced tests considered by 4th grade language arts standard. An asterisk indicates that the above criteria were met for a standard to be adequately measured by the items on the test.

Table 1. Alignment of norm-referenced tests to 4th grade language arts standards.

	Norm-referenced tests				
	CAT	MAT	SAT	TN	ITBS
Reading Standard:					
1	*	*	*	*	*
2	*	*	*	*	*
3	*	*	*	*	*
4	*	*	*		*
5					
6	*	*			
7					
8					
Writing Standard:					
1	*	*	*	*	*
2					
3					*
4					
5					
Speaking Standard:					
1					
2					
Listening Standard:					
1			*		

As shown in Table 1, the first three reading and first writing standards were judged by the teachers to be adequately measured by the five norm-referenced tests. These standards ask students to demonstrate basic knowledge and skills in reading comprehension, vocabulary, reading strategies, and writing mechanics. Reading standard 4 addresses content related to reference material knowledge and skills and was judged by the teachers to be adequately measured by four of the five tests. Three standards, reading

standard 6, writing standard 4, and listening standard 1 were judged by teachers to be adequately measured by two or fewer tests.

8th Grade Language Arts

Table 2 shows the results for the alignment of the five norm-referenced tests considered by 8th grade language arts standard. An asterisk indicates that the above criteria were met for a standard to be adequately measured by the items on the test.

Table 2. Alignment of norm-referenced tests to 8th grade language arts standards.

	Norm-referenced tests				
	CAT	MAT	SAT	TN	ITBS
Reading Standard:					
1	*	*	*	*	*
2	*	*	*		*
3	*				
4					
5					
6					
7			*	*	*
Writing Standard:					
1	*	*	*	*	*
2	*	*			*
3	*	*	*	*	*
4					
5	*				
Speaking Standard:					
1					
2					
Listening Standard:					
1			*		
2			*		

As shown in Table 2, the first reading and the first and third writing standards were judged by the teachers to be adequately measured by the five norm-referenced tests. These standards ask students to demonstrate basic knowledge and skills in reading comprehension, writing conventions, and editing. Reading standard 2 addresses content

related to reference material knowledge and skills and was judged by the teachers to be adequately measured by four of the five tests. Two standards, reading standard 7 and writing standard 2 were judged by teachers to be adequately measured by three of the five tests. Four standards, reading standard 3, writing standard 5, and listening standards 1 and 2 were judged by teachers to be adequately measured by two or fewer tests.

11th Grade Language Arts

Table 3 shows the results for the alignment of the five norm-referenced tests considered by 11th grade language arts standard. An asterisk indicates that the above criteria were met for a standard to be adequately measured by the items on the test.

Table 3. Alignment of norm-referenced tests to 11th grade language arts standards.

	Norm-referenced tests				
	CAT	MAT	SAT	TN	ITED
Reading Standard:					
1	*	*	*	*	*
2	*	*	*	*	*
3	*				*
4					
5					*
6			*		
7	*			*	
8	*				*
Writing Standard:					
1	*	*	*	*	*
2	*	*	*	*	*
3	*	*	*	*	*
4					
5	*				
Speaking Standard:					
1					
2					
Listening Standard:					
1					

As shown in Table 3, the first two reading and the first three writing standards were judged by the teachers to be adequately measured by the five norm-referenced tests. These standards ask students to demonstrate basic knowledge and skills in reading comprehension, reference materials, writing conventions, organization, and editing. Six standards, reading standards 3, 5, 6, 7, 8, and writing standard 5 were judged by teachers to be adequately measured by two or fewer tests.

Conclusion

Many states are adopting standards-led assessment programs that utilize commercially available norm-referenced tests. Our findings suggest that this may be an incomplete and inadequate approach for collecting or reporting student performance data. Further, depending on the state's articulated content standards, the test they select may align better or worse than another commercially available test. Gaps in what these norm-referenced tests measure exist across grade. Finally, the tests themselves are not interchangeable as each one has unique characteristics that align better or worse with the language arts standards.

References

California Achievement Tests (5th ed.). (1993). CTB Macmillan/McGraw Hill.

Iowa Tests of Basic Skills (Form M). (1996). Riverside Publishing Company (Houghton Mifflin Company).

Linn, R.L & Herman, J.L. (1997). *A Policymaker's Guide to Standards-Led Assessment*. Denver, CO: Education Commission of the States & National Center for Research on Evaluation, Standards, and Student Testing.

Metropolitan Achievement Tests (7th ed.). (1993). The Psychological Corporation.

Schmidt, W.H. (1999). *Using off-the-shelf tests in large-scale assessment systems*. Presentation at the National Conference on Large-Scale Assessment. Snowbird, UT.

Stanford Achievement Test (9th ed.). (1997). Harcourt Brace & Company.

TerraNova. (1997). CTB/McGraw Hill.

Webb, N.L. & Smithson, J. (1999, April). *Alignment between standards and assessments in mathematics for grade 8 in one state*. Paper presented at the annual meeting of the American Educational Research Association. Montreal, Quebec.

Webb, N.L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6. Madison, WI: University of Wisconsin – Madison.

Wixson, K.K., Fisk, M.C., Dutro, E., & McDaniel, J. (1999, June). *The alignment of state standards and assessments in elementary reading*. Presentation at the National Conference on Large-Scale Assessment. Snowbird, UT.

BEST COPY AVAILABLE